

# Genotyping of turkey coronavirus field isolates from various geographic locations in the United States based on the spike gene

Yi-Ning Chen<sup>1,2</sup> · Chien Chang Loa<sup>1,3</sup> · Mustafa Mohammed-Khair Ababneh<sup>1,4</sup> ·  
Ching Ching Wu<sup>1,5</sup> · Tsang Long Lin<sup>1</sup>

Received: 2 April 2015 / Accepted: 27 July 2015  
© Springer-Verlag Wien 2015

**Abstract** Turkey flocks have experienced turkey coronavirus enteritis sporadically in the United States since the 1990s. Twenty-four field isolates of turkey coronavirus (TCoV) from multiple states in the United States were recovered from 1994 to 2010 to determine the genetic relationships among them. The entire spike (S) gene of each TCoV isolate was amplified and sequenced. Pairwise comparisons were performed using the Clustal W program, revealing 90.0 % to 98.4 % sequence identity in the full-length S protein, 77.6 % to 96.6 % in the amino terminus of the S1 subunit (containing one hypervariable region in S1a), and 92.1 % to 99.3 % in the S2 subunit at the deduced amino acid sequence level. The conserved motifs, including two cleavage recognition sequences of the S protein, two heptad repeats, the transmembrane domain,

and the Golgi retention signal were identified in all TCoV isolates. Phylogenetic analysis based on the full-length S gene was used to distinguish North American TCoV isolates from French TCoV isolates. Among the North American TCoV isolates, three distinct genetic groups with 100 % bootstrap support were observed. North Carolina isolates formed group I, Texas isolates formed group II, and Minnesota isolates formed Group III. The S genes of 24 TCoV isolates from the United States remained conserved because they contained predominantly synonymous substitutions. The findings of the present study suggest endemic circulation of distinct TCoV genotypes in different geographic locations.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00705-015-2556-2) contains supplementary material, which is available to authorized users.

✉ Tsang Long Lin  
tllin@purdue.edu

<sup>1</sup> Department of Comparative Pathobiology, Purdue University, 406 South University Street, West Lafayette, IN 47907, USA

<sup>2</sup> Department of Bioscience Technology, Chung Yuan Christian University, 200 Chung Pei Road, Chung Li 32023, Taiwan, ROC

<sup>3</sup> Present Address: Mylan Pharmaceuticals Inc., 3711 Collins Ferry Road, Morgantown, WV 26505, USA

<sup>4</sup> Present Address: Department of Basic Veterinary Medical Sciences, Jordan University of Science and Technology, Irbid 22110, Jordan

<sup>5</sup> Present Address: School of Veterinary Medicine, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC

## Introduction

Turkey coronavirus enteritis of varying severity, caused by turkey coronavirus (TCoV), has been reported in turkey flocks from multiple states in the United States since the 1990s. The major clinical signs of TCoV infection include depression, ruffled feathers, diarrhea, decreased body weight, and uneven flock growth. The most apparent gross lesions are markedly distended intestines with gaseous and watery content, particularly in the ileum and ceca. Salient histopathologic findings include shortening of the intestinal villi, an increase in crypt depth, and widening of intervillous spaces [1]. When turkeys are infected with TCoV and other infectious agents such as astrovirus, small round virus, and *Escherichia coli* (*E. coli*), they can develop poult enteritis-mortality syndrome (PEMS), which causes high mortality [2, 3]. Subsequent experimental studies of the TCoV isolates VR-911 and TCoV/ON/MG10/08 from Canada have shown that TCoV can cause symptoms

similar to those caused by PEMS [3, 4]. Therefore, TCoV has been suggested to be the major causative pathogen for turkey enteritis, and secondary infections caused by other opportunistic microorganisms enhance the severity of TCoV enteritis and contribute to the development of PEMS. Turkey enteritis associated with TCoV infection has caused substantial economic losses in Indiana, North Carolina, Arkansas, and other states in the United States [5, 6], as well as in Canada [4], Europe [2, 7], and Brazil [8]. Currently, there are no vaccines to prevent the disease, and treatment of infected turkeys is often unsuccessful.

A member of the species *Avian coronavirus* (CoV) in the genus *Gammacoronavirus* and family *Coronaviridae*, TCoV has a positive single-stranded RNA genome that is approximately 27 Kb in size. The major structural proteins of TCoV include the spike (S), envelope (E), matrix (M), and nucleocapsid (N) proteins. Comparisons of 3'-end coding regions [9, 10] as well as the full genomes [11] of TCoV isolates and infectious bronchitis virus (IBV) have suggested that TCoV arises through recombination in the S gene, because pairwise comparisons of S gene sequences have revealed only a 34 % similarity between TCoV isolates and IBV strains, whereas gene 3, M gene, gene 5, and N gene sequences have over 80 % similarity [9, 11, 12]. The S gene sequences of different TCoV isolates (93 %–99.7 %) are more conserved than those of various IBV strains (67.4 %–94.4 %), which could explain the close antigenic relationship of TCoV isolates compared with the distant antigenicity of different IBV serotypes [5, 10, 13, 14]. Investigations during TCoV outbreaks and genomic analyses of TCoV isolates have revealed that distinct TCoV isolates tend to circulate endemically, and their respective sequences group phylogenetically according to their state of origin [6, 7, 11]. However, these observations may have been biased because of the small number of TCoV sequences that were analyzed. In the present study, 24 TCoV isolates were recovered from clinical cases submitted to the Indiana Animal Disease Diagnostic Laboratory at Purdue University by turkey farms in Minnesota, Indiana, North Carolina, Missouri, Arkansas, Texas, South Carolina, and Pennsylvania between 1994 and 2010. The objective of the present study was to elucidate the relationship between the genotypes and geographic distribution of TCoV isolates from turkey farms in multiple states in the United States by using sequence analysis and comparing the full-length S gene.

## Materials and methods

### Clinical samples and virus purification

Twenty-four field isolates of TCoV were recovered from clinical cases submitted to the Animal Disease Diagnostic Laboratory at Purdue University by turkey farms in

Minnesota, Indiana, North Carolina, Missouri, Arkansas, Texas, South Carolina, and Pennsylvania between 1994 and 2010 (Table 1). Field cases of TCoV were confirmed by clinical signs, gross lesions, histopathologic findings, immunofluorescence antibody (IFA) assay with antiserum against TCoV/IN/540/94, electron microscopy, and reverse transcription polymerase chain reaction (RT-PCR). All 24 TCoV isolates were propagated five times in embryonated turkey eggs as described previously [1]. In brief, intestines from TCoV-infected turkeys were homogenized as 20 % suspensions in chilled sterile phosphate-buffered saline and clarified by centrifugation at 3000 rpm for 10 minutes at 4 °C. The supernatant was filtered through a 0.22-μm membrane filter (Millipore, Bedford, MA, USA). The filtrate was inoculated into the amniotic cavity of 22-day-old embryonated turkey eggs. The embryo intestines were harvested after 3 days of incubation for virus purification. The harvested intestines were homogenized and clarified at 3000 rpm for 10 minutes at 4 °C. The supernatant was layered on top of 30 % and 60 % sucrose and clarified using ultracentrifugation in an SW28 rotor at 24,000 rpm for 3 hours at 4 °C in an Optima XL-100K ultracentrifuge (Beckman Coulter, Fullerton, CA, USA). The interface between 30 % and 60 % sucrose was collected and placed on top of a continuous 40 %–60 % sucrose gradient and clarified by ultracentrifugation at 24,000 rpm for 20 hours at 4 °C. A band of buoyant density 1.16–1.24 g/mL (containing TCoV) was collected and saved at -80 °C as the viral stock.

### RNA and cDNA

The viral RNA was extracted from the purified virus using RNApure™ Reagent (GenHunter, Nashville, TN, USA) and chloroform, followed by precipitation using cold isopropyl alcohol and ethanol. The extracted RNA was reverse transcribed to cDNA using SuperScript™ III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The first reaction was 10 minutes of incubation with the RNA, random hexamer primer (100 ng/μL), and 10 mM dNTPs at 70 °C, followed by 1 minute on ice. The second reaction was 5 minutes of incubation with a mixture of 5x first-strand buffer, 0.1 M of dithiothreitol (DTT), 1 U of SuperScript™ III reverse transcriptase, and 40 U of RNaseOUT™ (Invitrogen, Carlsbad, CA, USA) at 25 °C, followed by 1 hour of incubation at 50 °C and 15 minutes of inactivation at 70 °C.

### PCR amplification and sequencing

The full-length S gene (approximately 3.9 kb) was amplified by PCR using the cDNA of each TCoV isolate with the

**Table 1** Turkey coronavirus (TCoV) isolates and other coronaviruses used in the molecular analysis of the spike (S) gene

Isolate <sup>a</sup>	Year	Location	Accession number	Reference
TCoV/MN/ATCC/76	1976	Minnesota, USA	EU022526	[15]
TCoV/MN/310/96	1996	Minnesota, USA	KF652237	This study
TCoV/IN/517/94	1994	Indiana, USA	GQ427175	[11]
TCoV/IN/517-Purdue/94	1994	Indiana, USA	KF652222	This study
TCoV/IN/540/94	1994	Indiana, USA	EU022525	[15]
TCoV/IN/671/04	2004	Indiana, USA	KF652220	This study
TCoV/IN/834/04	2004	Indiana, USA	KF652221	This study
TCoV/IN/287/09	2009	Indiana, USA	KF652219	This study
TCoV/IN/421/10	2010	Indiana, USA	KF652218	This study
TCoV/NC/1020/96	1996	North Carolina, USA	KF652227	This study
TCoV/NC/DK9/97	1997	North Carolina, USA	KF652240	This study
TCoV/NC/1440/99	1999	North Carolina, USA	KF652238	This study
TCoV/NC/183-1/02	2002	North Carolina, USA	KF652230	This study
TCoV/NC/625/02	2002	North Carolina, USA	KF652224	This study
TCoV/NC/20/09	2009	North Carolina, USA	KF652223	This study
TCoV/NC/1741/12	2012	North Carolina, USA	KJ146015	[6]
TCoV/NC/1729/12	2012	North Carolina, USA	KJ146017	[6]
TCoV/NC/1728/12	2012	North Carolina, USA	KJ146018	[6]
TCoV/MO/168/99	1999	Missouri, USA	KF652229	This study
TCoV/MO/2216/99	1999	Missouri, USA	KF652231	This study
TCoV/MO/2580/99	1999	Missouri, USA	KF652232	This study
TCoV/MO/333/06	2006	Missouri, USA	KF652239	This study
TCoV/AR/428/96	1996	Arkansas, USA	KF652234	This study
TCoV/AR/1425/96	1996	Arkansas, USA	KF652228	This study
TCoV/AR/1811/12	2012	Arkansas, USA	KJ146016	[6]
TCoV/AR/1913/12	2012	Arkansas, USA	KJ146019	[6]
TCoV/AR/1744/12	2012	Arkansas, USA	KJ146020	[6]
TCoV/TX/1038-Purdue/98	1998	Texas, USA	KF652225	This study
TCoV/TX/1038/98	1998	Texas, USA	GQ427176	[11]
TCoV/TX/GL/01	2001	Texas, USA	GQ427174	[11]
TCoV/TX/176-2/02	2002	Texas, USA	KF652226	This study
TCoV/SC/284/96	1996	South Carolina, USA	KF652233	This study
TCoV/PA/682/98	1998	Pennsylvania, USA	KF652236	This study
TCoV/PA/640/02	2002	Pennsylvania, USA	KF652235	This study
TCoV/VA/74/03	2003	Virginia, USA	GQ427173	[11]
TCoV/ON/MG10/08	2008	Ontario, Canada	EU095850	[4]
TCoV/FR070341j/07	2007	France	GQ411201	[7]
TCoV/FR080147c/08	2008	France	FN434203	[7]
TCoV/FR080183j/08	2008	France	FN545819	[7]
TCoV-0/BR/08	2008	Brazil	HQ853282	Direct submission
IBV/Beaudette CK	1937	United Kingdom	AJ311317	[23]
IBV/Arkansas DPI	1981	Arkansas, USA	GQ504720	[24]
Guinea fowl CoV/FR/2011	2011	France	HF544506	[25]
Beluga whale SW1 CoV	2008	Missouri, USA	EU111742	[26]

primers Sup and Sdown3 (Online Resource 1). The mixture (64:1, v:v) of Taq (Promega Corp., Madison, WI, USA) and Pfu DNA polymerases (Stratagene, La Jolla, CA, USA) with proofreading ability was used in a 96-well thermal

cycler (GeneAmp, Perkin-Elmer Cetus Corp., Norwalk, CT, USA) to maintain the fidelity of the PCR [9]. The PCR products were electrophoresed on 1 % agarose gels and purified using a Zymoclean™ Gel DNA Recovery Kit

(Zymo, Irvine, CA, USA) for further sequencing. Several primers (Online Resource 1) designed to sequence overlapping fragments covering the full length of the S gene of TCoV were used to determine the nucleotide sequences of the purified PCR products at the Purdue University Genomics Core Facility (West Lafayette, IN, USA). In addition, the purified PCR product was cloned into the PCR-II plasmid vector and used to transform *E. coli* strain TOP10F' according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA).

## Sequence analysis

Twenty-four TCoV isolates collected from eight states in the United States from 1994 to 2010 were purified and sequenced in our laboratory. TCoV/IN/549/94 was analyzed in a previous study [15], and the full-length S gene sequences of the other 23 TCoV isolates were published for the first time in the present study (Table 1). The putative peptide cleavage site separating the amino-terminus of the S1 subunit from the carboxyl terminus of the S2 subunits as well as a second possible peptide cleavage site in the S2 subunit were detected using the ProP server (<http://www.cbs.dtu.dk/services/ProP/>). The S1 subunit of TCoV was further designated as S1a at the amino-terminus (1-204 in TCoV/IN/540/94) and S1b at the carboxyl-terminus (205-536 in TCoV/IN/540/94) [7]. Figure 1 shows a diagram of the S protein of TCoV. The nucleotide and deduced amino acid sequence similarities of the S genes of all 24 TCoV isolates were analyzed using the Clustal W alignment method in MEGA6 [16]. A phylogenetic tree based on full-length nucleotide sequences of the S gene was constructed using the maximum-likelihood method and the Kimura 2-parameter model. A phylogenetic tree based on deduced

amino acid sequences of the S1a subunit, containing a hypervariable region (HVR), was constructed using the neighbor-joining method and the Jones–Taylor–Thornton model. A codon-based Z-test of positive selection for S gene sequences of various TCoV isolates was conducted to analyze the differences in the number of nonsynonymous (dN) and synonymous (dS) substitutions per site by using the Nei–Gojobori method [17] in MEGA6. The variance of both trees and codon-based Z-test were validated using 1000 bootstrap replicates.

## Nucleotide sequence accession numbers

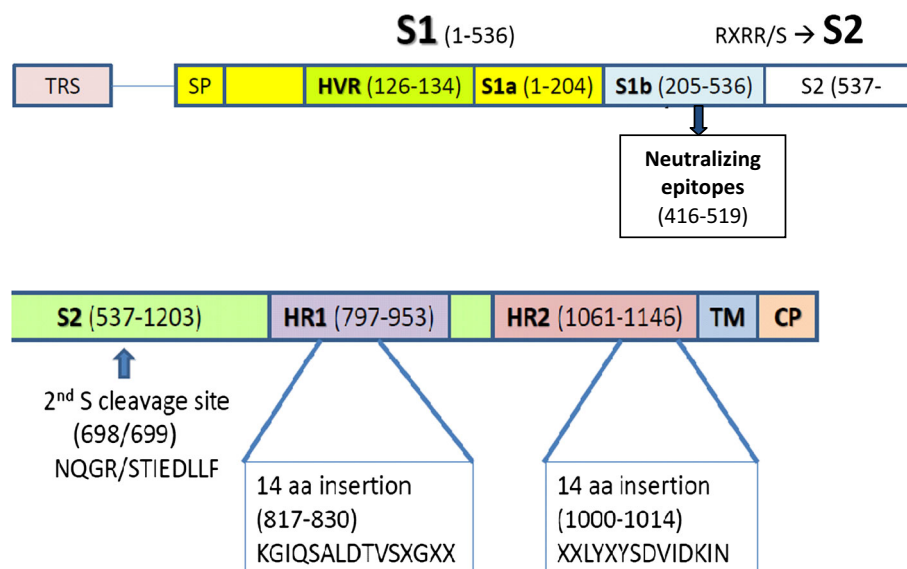
The S sequences of TCoV isolates reported in the present study were submitted to the GenBank database, and their accession numbers ranged from KF652218 to KF652240. The accession numbers of other CoVs used for phylogenetic analysis are also listed in Table 1.

## Results

### Genetic analysis of the spike sequences of turkey coronavirus isolates

The sizes of the S genes of TCoV isolates reported in the present study ranged from 3609 to 3630 nucleotides. All 24 TCoV isolates exhibited similar S protein sequences (Fig. 1). The consensus transcription-regulating sequence (TRS), CTGAACAA, was identified 52 nucleotides upstream of the start codon of the S protein. Several conserved motifs and one HVR were found in all TCoV isolates, and their sequences are listed for comparison in Table 2. The consensus motif RXRR/X (X is any amino

**Fig. 1** Schematic diagram of the spike protein of turkey coronavirus. TRS, transcription regulatory sequence; SP, signal peptide; HVR, hypervariable region; HR, heptad repeat; TM, transmembrane domain; CP, cytoplasmic peptides. Numbers in parentheses indicate the amino acid position from the start codon of the spike protein from TCoV/IN/540/94 (EU022525)



**Table 2** Conserved motifs in the spike (S) protein cleavage sites and the sequences of the hypervariable region (HVR) of turkey coronavirus (TCoV) isolates

Isolate	S full length (aa)	Hypervariable region (HVR) (126-134)*	S1/S2 cleavage (536)*	S2 2nd site (698)*
TCoV/ATCC/MN/76	1203	LSMGN—ST	RRRR ATS	NQGR/S
TCoV/310/MN/96	1203	LSMGN—ST	RRRR ATS	NQGR/S
TCoV/IN/517/94	1226	LATNVGKVTPP-ND	RKRR/S	NQGR/S
TCoV/IN/517-Purdue/94	1205	LGT—KVG D—HD	RERR/S	NQGR/S
TCoV/IN/540/94	1203	LGT—KVG D—HD	RERR/S	NQGR/S
TCoV/IN/671/04	1209	IDTNVGKVTPP-HE	RKRR/S	NQGR/S
TCoV/IN/834/04	1209	VYTNVGKVTPP-NE	RKRR/S	NQGR/S
TCoV/IN/287/09	1209	LITNVGD—HD	RKRR/S	NQGR/S
TCoV/IN/421/10	1208	IETNVGK—TPP-ND	RRRR/S	NQGR/S
TCoV/NC/1020/96	1208	LVTNAGNVKP—DD	RKRR/S	NQGR/S
TCoV/NC/DK9/97	1209	LHTNVGKVTPP-ND	RKRR/S	NQGR/S
TCoV/NC/1440/99	1207	VSTNAGKVR—ND	RERR/S	NQGR/S
TCoV/NC/183-1/02	1208	IGTNVGKKPP—NE	RKRR/S	NQGR/S
TCoV/NC/625/02	1208	VATNVGKSAPWAND	RRRR/S	NQGR/S
TCoV/NC/20/09	1209	INTNAGKITLP-NE	RKRR/S	NQGR/S
TCoV/NC/1728/12	NA <sup>b</sup>	MDTNVGK—PN-ND	RRRR/S	NA <sup>b</sup>
TCoV/NC/1729/12	NA <sup>b</sup>	IDTNVGK—PN-ND	RRRR/S	NA <sup>b</sup>
TCoV/NC/1741/12	NA <sup>b</sup>	IDTNVGK—PN-NH	RRRR/S	NA <sup>b</sup>
TCoV/MO/168/99	1207	VGTNAGKQ—KE	RKRR/S	NQGR/S
TCoV/MO/2216/99	1206	VSTNAGQKP—NE	RKRR/S	NQGG/S
TCoV/MO/2580/99	1206	IGTNAGKV—ND	RKRR/S	NQGR/S
TCoV/MO/333/06	1206	IDTNAGKVGP—KD	RKRR/S	NQGR/S
TCoV/AK/428/96	1203	LKVAN—SH—	RKRR/S	NQGR/S
TCoV/AK/1425/96	1204	IETNVGK—SH—NE	RKRR/S	NQGR/S
TCoV/AK/1744/12	NA <sup>b</sup>	VGTNAGKVSAPNE	RKRR/S	NA <sup>b</sup>
TCoV/AK/1913/12	NA <sup>b</sup>	IDTNVGK—S—YD	RKRR/S	NA <sup>b</sup>
TCoV/TX/1038-Purdue/98	1202	INTKG—ND	RMRR/S	NQGR/S
TCoV/TX/1038/98	1223	IGTNADKKS—ND	RTRR/S	NQGR/S
TCoV/TX/GL/01	1225	IEINAGEVT—ND	RTRR/S	NQGR/S
TCoV/TX/176-2/02	1207	IGTNADKKS—ND	RMRR/S	NQGR/S
TCoV/SC/284/96	1207	LVTNVGNVKT—AD	RKRR/S	NQGR/S
TCoV/PA/682/98	1203	LSMGN—ST	RRRW ATS	NQGR/S
TCoV/PA/640/02	1208	IYTNVGKLTP—ND	RKRR/S	NQGR/S
TCoV/VA/74/03	1226	LATNVGKVTPPND	RKRR/S	NQGR/S
TCoV/ON/MG10/08	1226	LATNVGKVTPPND	RKRR/S	NQGR/S
TCoV-0 ~ 10/BR/08		LATNVGKVTPPND	RKRR/S	NQGR/S
TCoV/FR070341j/07	1198	SIGNFVLNS	RSRR/S	PQGR/S
TCoV/FR080147c/08	1198	SIGNFVLNS	RSRR/S	PQGR/S
TCoV/FR080183j/08	1198	SIGNFVLNS	RSRR/S	PQGR/S

<sup>a</sup> The positions of HVR, S1/S2 cleavage, and S2 second predicted cleavage site in parentheses were relative to the start codon of the spike protein of TCoV/IN/540/94 (EU022525)

<sup>b</sup> NA: data not available

acid, R is arginine, and slash [/] indicates the cleavage position) was found at the cleavage site for the S1 and S2 subunits in 22 TCoV isolates analyzed in the present study and the French TCoV isolates, except for the isolates TCoV/MN/310/96 and TCoV/PA/682/98, which had the

amino acid sequence ATS followed by the cleavage site, similar to TCoV/MN/ATCC/76. A conserved sequence, NQGR/S, resembling the furin-dependent cleavage site in IBV [18] was identified in the S2 subunit in all TCoV isolates except the isolate TCoV/MO/2216/99, in which the



critical arginine was mutated to glycine and the second probable protein cleavage site was lost. Rather than NQGR/S, French TCoV isolates had the PQGR/S sequence as the conserved cleavage motif in the S2 subunit. Only one HVR, spanning amino acid positions 126 to 134 (TCoV/IN/540/94) was found in the TCoV isolates, rather than the three HVRs identified in IBV [7]. Two 14-amino-acid insertions in heptad repeats (HR1 and HR2), the consensus motif (YIKWPWYVWL) in the transmembrane domain, and the late Golgi retention signal (YYTTF) for S protein were also observed in all 24 TCoV isolates. Among the 45 amino acid residues of the neutralizing-epitope-containing S fragment in the S1 subunit identified in a previous study [19], 33 consensus residues were observed among the 24 TCoV isolates (Online Resource 2).

### Comparison of the spike sequences of turkey coronavirus isolates

A pairwise comparison of the deduced amino acid sequences of the 24 TCoV isolates showed that the sequence identity ranged from 90.0 % to 98.4 % for the full-length S protein, 77.6 % to 96.6 % for the S1a subunit containing HVR, and 92.1 % to 99.3 % for the S2 subunit (Online Resource 3). No positive selection for the S gene was observed among the TCoV isolates (Online Resource 4). The values of  $d_S$  were greater than those of  $d_N$  in all comparisons except three pairs of TCoV isolates (TCoV/IN/540/94 and IN/517-Purdue/94, IN/671/04 and IN/834/04, MO/2216/99 and MO/168/99), which had similar  $d_S$  and  $d_N$  values and shared high sequence identity in the S gene, exceeding 95 %.

### Phylogenetic analysis of the spike gene

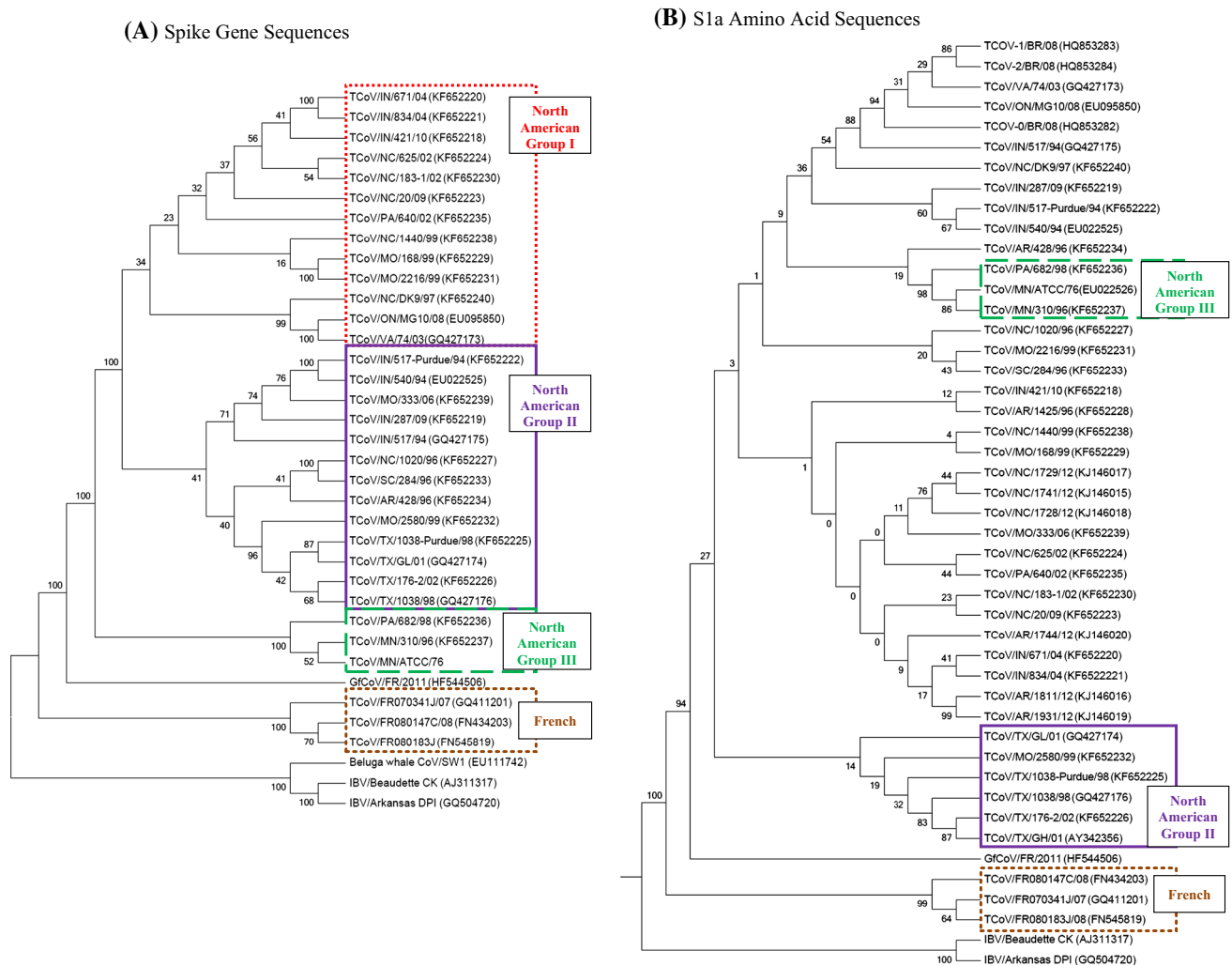
Phylogenetic trees based on the full-length S nucleotide sequences (Fig. 2A) and S1a amino acid sequences containing HVRs (Fig. 2B) of different CoVs of the genus *Gammacoronavirus* were generated. As shown in Figures 2A and 2B, the IBV strains were separated from TCoV isolates, and North American TCoV isolates were separated from French TCoV isolates. Three genetic groups, referred to as groups I, II, and III, were observed in North American TCoV isolates (Fig. 2A). Group I included all North Carolina isolates except TCoV/NC/1020/96, Indiana isolates from 2004 and 2010, Missouri isolates from 1999, and isolates from Pennsylvania, Virginia, and Canada. Group II contained all Texas isolates, Indiana isolates from 1994 and 2009, one Missouri isolate each from 1999 and 2006, and North Carolina, South Carolina, and Arkansas isolates from 1996. Group III was composed of two Minnesota isolates that were isolated 20 years apart and TCoV/PA/682/98, and they had sequence identities

higher than 98.5 %. Because of the high degree of variation, most phylogenetic groupings based on the S1a deduced amino acid sequences did not have a bootstrap value over 50 % (Fig. 2B). Nevertheless, the Texas TCoV isolates of group II and all three TCoV isolates of group III shown in the phylogenetic tree based on the full-length S nucleotide sequences still clustered according to their S1a amino acid sequences containing their HVR.

### Discussion

Turkey coronavirus isolates from different geographic areas in the United States have been shown to be antigenically related to one another [5]. In the present study, an antiserum against the isolate TCoV/IN/540/94 reacted with all 24 TCoV isolates from TCoV-infected turkeys and embryos by IFA assay (data not shown). The close antigenicity among TCoV isolates was associated with the high similarity of the S gene sequences, which ranged from 90.4 % to 99.4 %. The S genes of TCoV isolates were conserved compared with the diverse S genes among various IBV strains, which range from 67.4 % to 94 %, resulting in the existence of many serotypes of IBV [10]. The emergence of new IBV serotypes has been postulated to involve the recombination of S genes of vaccine strains of IBV in the field [20, 21]. In a previous study, different serotypes of TCoV/VA/73/03, TX/1038/98, and IN/517/94 were identified by using a neutralization test in conjunction with real-time RT-PCR despite the high level of amino acid sequence identity (96 % to 98 %) among these TCoV isolates [11]. Additional studies are necessary to clarify the antigenic relationships among the various TCoV isolates and serotypes.

In the present study, similar to previous findings with IBV strains, most of the variations in the S protein sequences among TCoV isolates were observed in the amino-terminal half. Along the alignment of these S protein sequences, the region of sequence with the most variation was between residues 126 and 134 of TCoV/IN/540/94 from the start codon of the S protein. Various deletions occurred in this region in different TCoV isolates. This region is in the vicinity of HVR II (residues 117 to 131) of the IBV S protein. HVR I (residues 56 to 69), II, and III (residues 250 to 365) of the IBV S protein are associated with three neutralizing epitopes. The sequences of these regions could be used for differential diagnosis of IBV serotypes [13, 22]. By contrast, similar regions of high variation corresponding to HVR I or III of IBV were not detected among the TCoV isolates examined. These differences illustrate why the S proteins of IBV strains are more diverse than those of TCoV isolates. Similar phylogenetic trees were constructed using the full-length S and



**Fig. 2** Consensus bootstrap phylogenetic tree based on the full-length spike (S) gene nucleotide sequences (A) and the S1a amino acid sequences (B) of turkey coronavirus (TCoV) isolates, infectious bronchitis virus (IBV) strains, guinea fowl (Gf) CoV, and beluga whale SW1 CoV (accession number in parentheses). The nucleotide

S1a proteins. Thus, genotyping TCoV field isolates based on the S1a sequence containing HVR rather than the whole S1 gene or full-length S gene is more practical.

The observation that TCoV isolates originating from the same state were closely clustered together in the phylogenetic tree suggested endemic circulation of distinct TCoV genotypes in various geographic locations. Endemic circulation of distinct TCoV genotypes in France and North American are recognized because French TCoV isolates share only 60 % amino acid sequence identity in the S protein with North American TCoV isolates [7]. Distinct sources of recombination promoting the emergence of TCoV in North America and Europe has been suggested [7, 11]. The groupings of North Carolina, Texas, and Minnesota TCoV isolates also support the theory of endemic TCoV genotypes. The

sequence tree was constructed by the maximum-likelihood method and the Kimura 2-parameter model, and the amino acid sequence tree was constructed by the neighbor-joining method and the Jones-Taylor-Thontom model in MEGA 6. The bootstrap values were calculated from 1000 trees

TCoV isolates from the outbreaks in Arkansas and North Carolina in 2012 also clustered geographically [6] and could be placed phylogenetically in group I in the present study. The 99.3 % amino acid sequence identity of the S proteins of two TCoV isolates recovered 20 years apart in Minnesota (TCoV/MN/ATCC/76 and TCoV/MN/310/96) implied that the TCoV isolate MN/ATCC/76 remained endemic and that no substantial genetic changes occurred over two decades. Conservation of TCoV isolates is also shown in the result that no positive selection of the S protein was found among the 24 TCoV isolates. Because Indiana isolates from 2004 and 2010 clustered in group I with most North Carolina isolates and Indiana isolates from 1994 and 2009 clustered in group II with Texas isolates, it is most likely that the turkey sources were the same for the turkey

farms in North Carolina and Indiana in 2004 and 2010, whereas the turkey sources were the same for Texas and Indiana in 1994 and 2009.

In conclusion, the relationship between TCoV genotypes and the geographic distribution of TCoV presented in the present study provides crucial information for the monitoring and control of diseases associated with TCoV infection in the United States.

**Acknowledgments** The authors thank the US Department of Agriculture for the financial support, and Dr. David Hermes from Perdue Farm (Thorntown, IN, USA) for the turkey eggs and turkey poult used in virus isolation and propagation. The authors also thank Mr. Tom Hooper and Ms. Danna Schrader for their help in virus isolation, and Ms. Jing Dai for her assistance in RT-PCR amplification.

## References

- Lin TL, Loa CC, Tsai SC, Wu CC, Bryan TA, Thacker HL, Hooper T, Schrader D (2002) Characterization of turkey coronavirus from turkey poult with acute enteritis. *Vet Microbiol* 84:179–186
- Cavanagh D, Mawditt K, Sharma M, Drury SE, Ainsworth HL, Britton P, Gough RE (2001) Detection of a coronavirus from turkey poult in Europe genetically related to infectious bronchitis virus of chickens. *Avian Pathol* 30:355–368
- Yu MW, Ismail MM, Qureshi MA, Dearth RN, Barnes HJ, Saif YM (2000) Viral agents associated with poult enteritis and mortality syndrome: the role of a small round virus and a turkey coronavirus. *Avian Dis* 44:297–304
- Gomaa MH, Yoo D, Ojick D, Barta JR (2009) Infection with a pathogenic turkey coronavirus isolate negatively affects growth performance and intestinal morphology of young turkey poult in Canada. *Avian Pathol* 38:279–286
- Lin TL, Loa CC, Wu CC, Bryan T, Hooper T, Schrader D (2002) Antigenic relationship of turkey coronavirus isolates from different geographic locations in the United States. *Avian Dis* 46:466–472
- Day JM, Gonder E, Jennings S, Rives D, Robbins K, Tilley B, Wooming B (2014) Investigating turkey enteric coronavirus circulating in the Southeastern United States and Arkansas during 2012 and 2013. *Avian Dis* 58:313–317
- Maurel S, Toquin D, Briand FX, Queguiner M, Allee C, Bertin J, Ravillion L, Retaux C, Turblin V, Morvan H, Eterradossi N (2011) First full-length sequences of the S gene of European isolates reveal further diversity among turkey coronaviruses. *Avian Pathol* 40:179–189
- Teixeira MC, Luvizotto MC, Ferrari HF, Mendes AR, da Silva SE, Cardoso TC (2007) Detection of turkey coronavirus in commercial turkey poult in Brazil. *Avian Pathol* 36:29–33
- Lin TL, Loa CC, Wu CC (2004) Complete sequences of 3' end coding region for structural protein genes of turkey coronavirus. *Virus Res* 106:61–70
- Loa CC, Wu CC, Lin TL (2006) Comparison of 3'-end encoding regions of turkey coronavirus isolates from Indiana, North Carolina, and Minnesota with chicken infectious bronchitis coronavirus strains. *Intervirology* 49:230–238
- Jackwood MW, Boynton TO, Hilt DA, McKinley ET, Kissinger JC, Paterson AH, Robertson J, Lemke C, McCall AW, Williams SM, Jackwood JW, Byrd LA (2009) Emergence of a group 3 coronavirus through recombination. *Virology* 398:98–108
- Hughes AL (2011) Recombinational histories of avian infectious bronchitis virus and turkey coronavirus. *Arch Virol* 156:1823–1829
- Wang CH, Huang YC (2000) Relationship between serotypes and genotypes based on the hypervariable region of the S1 gene of infectious bronchitis virus. *Arch Virol* 145:291–300
- Jia W, Mondal SP, Naqi SA (2002) Genetic and antigenic diversity in avian infectious bronchitis virus isolates of the 1940s. *Avian Dis* 46:437–441
- Cao J, Wu CC, Lin TL (2008) Complete nucleotide sequence of polyprotein protein 1 and genome organization of turkey coronavirus. *Virus Res* 136:43–49
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Yamada Y, Liu XB, Fang SG, Tay FP, Liu DX (2009) Acquisition of cell-cell fusion activity by amino acid substitutions in spike protein determines the infectivity of a coronavirus in cultured cells. *PLoS one* 4:e6130
- Chen YN, Wu CC, Lin TL (2011) Identification and characterization of a neutralizing-epitope-containing spike protein fragment in turkey coronavirus. *Arch Virol* 156:1525–1535
- Mondal SP, Cardona CJ (2007) Genotypic and phenotypic characterization of the California 99 (Cal99) variant of infectious bronchitis virus. *Virus Genes* 34:327–341
- Ndegwa EN, Toro H, van Santen VL (2014) Comparison of vaccine subpopulation selection, viral loads, vaccine virus persistence in trachea and cloaca, and mucosal antibody responses after vaccination with two different Arkansas Delmarva Poultry Industry-derived infectious bronchitis virus vaccines. *Avian Dis* 58:102–110
- Lee C-W, Hilt DA, Jackwood MW (2003) Typing of field isolates of infectious bronchitis virus based on the sequence of the hypervariable region in the S1 gene. *J Vet Diagn Invest* 15:344–348
- Casais R, Thiel V, Siddell SG, Cavanagh D, Britton P (2001) Reverse genetics system for the avian coronavirus infectious bronchitis virus. *J Virol* 75:12359–12369
- Phillips JE, Jackwood MW, McKinley ET, Thor SW, Hilt DA, Acevedol ND, Williams SM, Kissinger JC, Paterson AH, Robertson JS, Lemke C (2012) Changes in nonstructural protein 3 are associated with attenuation in avian coronavirus infectious bronchitis virus. *Virus Genes* 44:63–74
- Liais E, Croville G, Mariette J, Delverdier M, Lucas MN, Klopp C, Lluch J, Donnadiou C, Guy JS, Corrand L, Ducatez MF, Guerin JL (2014) Novel avian coronavirus and fulminating disease in Guinea fowl, France. *Emerg Infect Dis* 20:105–108
- Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D (2008) Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J Virol* 82:5084–5088